DOCUMENT RESUME

ED 423 294 TM 029 101

AUTHOR Saxe, Geoffrey B.; Franke, Megan L.; Gearhart, Maryl; Howard, Sharon; Crockett, Michele

TITLE Teachers' Shifting Assessment Practices in the Context of Educational Reform in Mathematics. CSE Technical Report 471.

INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.

REPORT NO CSE-TR-471

PUB DATE 1997-12-00

NOTE 54p.

CONTRACT R305B60002

PUB TYPE Reports - Research (143) -- Tests/Questionnaires (160)

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS Educational Assessment; *Educational Change; Elementary Secondary Education; *Mathematics Teachers; *Performance Based Assessment; *Student Evaluation; Surveys; Test Construction; *Test Use

IDENTIFIERS *Reform Efforts

ABSTRACT

This paper presents a study of primary and secondary mathematics teachers' changing assessment practices in the context of policy, stakeholder, and personal presses for change. Two cohorts of teachers, one of 35 elementary, middle school, and high school teachers and the other of 24 elementary and middle school teachers participated. Using survey and interviews, teachers' reports of their uses of three forms of assessment were collected. Two forms were linked to reforms in mathematics education (open ended problems and rubrics) and one was linked to traditional practice (exercises). Findings reveal several trajectories of change in the interplay between assessment forms and the functions that they serve. Teachers may implement new assessment forms in ways that serve "old" functions; teachers may re-purpose "old" assessment forms in ways that reveal students' mathematical thinking. The developmental framework in this study provides a way to understand the dynamics of teacher development in relation to ongoing educational reforms. Three appendixes present the teacher survey, interview schedule, and teacher questionnaire. (Contains 4 figures and 12 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 294

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

Teachers' Shifting Assessment Practices in the Context of Educational Reform in Mathematics

CSE Technical Report 471

Geoffrey B. Saxe, Megan L. Franke,
Maryl Gearhart, Sharon Howard, and Michele Crockett
CRESST/University of California, Los Angeles

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

▸ **UCLA Center for the Study of Evaluation**

In collaboration with:

- **University of Colorado**
- **NORC, University of Chicago**
- **LRDC, University of Pittsburgh**
- **University of California, Santa Barbara**
- **University of Southern California**
- **The RAND Corporation**

TM029101

BEST COPY AVAILABLE

Teachers' Shifting Assessment Practices in the Context of Educational Reform in Mathematics

CSE Technical Report 471

Geoffrey B. Saxe, Megan L. Franke,
Maryl Gearhart, Sharon Howard, and Michele Crockett
CRESST/University of California, Los Angeles

December 1997

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 1998 The Regents of the University of California.

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

TEACHERS' SHIFTING ASSESSMENT PRACTICES IN THE CONTEXT OF EDUCATIONAL REFORM IN MATHEMATICS

Geoffrey B. Saxe, Megan L. Franke,
Maryl Gearhart, Sharon Howard, and Michele Crockett
CRESST/University of California, Los Angeles

Abstract

This paper presents a study of primary and secondary mathematics teachers' changing assessment practices in the context of policy, stakeholder, and personal presses for change. Using survey and interviews, we collected teachers' reports of their uses of three forms of assessment, one linked to traditional practice (exercises), and two linked to reforms in mathematics education (open ended problems and rubrics). Findings revealed several trajectories of change in the interplay between assessment forms and the functions that they serve. Teachers may implement new assessment form in ways that serve 'old' functions; teachers may re-purpose 'old' assessment forms in ways that reveal students' mathematical thinking. Our developmental framework provides a way to understand the dynamics of teacher development in relation to ongoing educational reforms.

The field of mathematics education has experienced waves of reform throughout its history, and each wave has been marked by challenges to teachers (Tyack & Cuban, 1995). In the recent climate of reform, particular value is placed on problem solving and conceptual understanding, a marked departure from the more traditional focus on accuracy and procedural skills (California State Department of Education, 1992; NCTM, 1993, 1995). New mathematics curriculum has been developed to engage students in problem solving, and new methods of assessment have been developed to evaluate the ways that students interpret problems and construct strategies for their solution. These new approaches, and the principles and mathematics that underlie them, are challenging to understand. Mathematics teachers are being pressed to implement these new approaches or to adapt their existing practices to fit the reform recommendations. We know that they are challenged, but we understand little of the pathways by which they develop competence with the new forms and functions of practice. Pressed to change, teachers shift in the character of their instructional and assessment practices in ways we do not yet understand (Goldsmith & Schifter, 1997; Nelson, 1997).

The purpose of the study we report here was to investigate patterns of change in K-12 mathematics teachers' methods of classroom assessment. The teachers participating in the study were engaged in a long-term professional development program, and thus they were receiving encouragement and support for their efforts to implement new forms of assessment and to use them to serve functions aligned with reform.

Framework

To guide our inquiry, we use a framework for conceptualizing patterns of development in teachers' assessment practices. We start with two assumptions. First, teachers construct and re-construct their assessment activities on a daily basis, sustaining a network of routines in classroom life as they adjust to or resist a matrix of policy, stakeholder, and personal presses for change. Second, we can understand development over time in teachers' assessment practices as an interplay between assessment forms and the assessment functions that these forms serve: In the context of presses, teachers may re-purpose forms of assessment to accomplish new assessment functions, and teachers also may adopt new assessment forms to serve prior assessment functions.

Presses

Teachers work in a complex profession in which they are pressed to change or maintain their ongoing practice in relation to a wide range of factors (Jones, 1997). We conceptualize these factors as consisting of three types. (1) Various presses at the *institutional level* are regarded as levers for change, meaning that they provide policy makers with means of supporting or inhibiting changes in teachers' practices. Such levers include standards set forth by professional and state organizations, curricular materials, district testing, and professional development programs. Depending upon the content of the standards, the nature of curricular materials, the content of the tests, or the strength and orientation of the support programs, these factors can press teachers towards implementing particular visions of instruction or assessment. (2) Local interactions with *key stakeholders*—parents, administrators, colleagues, and the students themselves—create unique presses of their own. Regular interactions with these stakeholders—some institutionalized, some informal—may create tensions and/or supports in interpreting and adapting policy to local circumstances and sometimes lead to local 'spins' on current policies. (3) Finally, *teachers* themselves create their own internal presses, interpreting their ongoing practices in terms of their own values about what constitutes meaningful and useful assessment activities (Fennema, Carpenter, Franke, & Carey, 1992; Shulman, 1987; Thompson, 1992; Wood, Cobb, & Yackel, 1991).

Teachers' Assessment Practices

Scribner & Cole's working definition of "practice" provides a useful framework for our focus on teachers' assessment practices as situated in a network of policy, stakeholder, and personal presses.

> [A practice is a] . . . recurrent, goal-directed sequence of activities using a particular technology and particular systems of knowledge. We use the term "skills" to refer to the coordinated sets of actions involved in applying this knowledge in particular settings. A practice, then consists of three components: technology, knowledge, and skills . . . [and] . . . refers to socially developed and patterned ways of using technology and knowledge to accomplish tasks. (Scribner & Cole, 1981, pp. 236)

Following Scribner & Cole, we conceptualize teachers' assessment practices in terms of the technologies, knowledge, and skills that are supported and

constrained by the institutional, stakeholder, and personal presses we noted above.

Technologies are symbolic or material forms often with prior histories and used to accomplish particular goals in practices. In the case of assessment practices, the technologies that we target are (a) assessment forms used for *eliciting performances* from students—such as exercises (short routine problems with a single correct solution) and open-ended problems (less routine problems with multiple strategies or solutions possible), and (b) assessment forms for *evaluating performances*, such as scores (percent correct, numerical tally of total correct) and rubrics (ordinal levels pointing to qualitative analysis of performance). An assessment, then, is a method of eliciting a performance and evaluating it, and thus it entails a coordination of two assessment forms.

The presses that support, constrain, or inhibit the availability and use of assessment forms are varied. They occur at the institutional level (states or districts may mandate, professional development programs may recommend), at the level of interested stakeholder groups (people push teachers either to try new things, or keep using the old ones), and at the personal level (teachers' interests in trying new assessment forms or satisfaction with prior ones).

In making use of a particular form of assessment whether for eliciting or evaluating performance, teachers draw upon their *knowledge* and beliefs about students' mathematics their knowledge of mathematics, and their knowledge of assessment. For example, some elementary teachers may know the procedures for solving computational problems, but have little understanding of the mathematical concepts underlying these procedures. In eliciting and evaluating students' developing competence with rational number operations and concepts, they may thus focus on what they know—adherence to procedures—rather than students' understanding of the mathematical rationale for the procedures. Further, even teachers with considerable knowledge of the subject matter may nevertheless have limited understanding of their students. They may believe that children either understand a given concept, or not, without recognizing the diversity of students' developing conceptual understandings. A wide range of factors may support, constrain, or inhibit teacher knowledge. Institutional presses include professional support and teachers' guides. Stakeholders may push teachers to acquire greater knowledge, while others may be invested in maintenance of the status quo. Teachers themselves may feel satisfied with their

current knowledge, or they may feel a need to learn more about assessment or children's mathematics.

Assessment *skills* refer to the actions involved in the implementation of assessment practices in classrooms. Teachers must learn to coordinate technologies for eliciting and evaluating complex performances. Various presses influence teachers' developing skills with assessment practices. Institutional factors include opportunity for assessment training (e.g., district scoring), professional support, and teachers' guides; key stakeholders may push teachers to acquire greater assessment skill or press them to maintain existing methods; finally, teachers build on their current skills in developing, refining, or maintaining their assessment practices.

Relations between Teachers' Practices and Presses on Practice

In response to presses, teachers adopt new assessment *forms* that are designed to serve new assessment *functions*. For example, teachers are asked to incorporate open-ended problems into their assessment activities (assessment form); such problems are intended to provide teachers the opportunity to gain insight into students' methods of problem solving and their understandings of mathematical concepts (assessment functions). For many teachers, the adoption of new *technologies* (new forms of assessment and new functions for these forms) requires new *knowledge* of the subject matter of mathematics and of frameworks that capture the sense that children make of the mathematics. Adoption also requires new *skills* that take time to develop, such as orchestrating lessons in ways that interweave assessment activities and instructional activities. Without such knowledge and skill, teachers will be unable to use the assessment forms to serve the functions promoted in reform.

Our Study

The purpose of our study was to document how mathematics teachers' methods of assessment shift over time in relation to the presses of institutions, stakeholders, and teachers' own efforts to change. Of particular interest were changes in the forms of assessment and the functions that they serve in teachers practices. We conducted the work in two phases.

In the first phase, we fielded surveys to K-12 teachers participating in a voluntary long-term professional development program. Representing a

diversity of schools and districts in Greater Los Angeles, these teachers shared in common an interest in working with a community of like-minded professionals to implement reforms in mathematics education. To capture the patterns of change, we asked the teachers to report on the frequency with which they were currently using various kinds of assessment forms for eliciting student performances (e.g., exercises, open-ended problems) as well as various forms of evaluation (e.g., percentage correct, rubric scores), and to compare their current uses with their uses in the past and their anticipated uses in the future. The survey responses provided us with evidence of patterns of change over time. To shed light on functions that the forms of assessment serve in teachers' practices, as well as how shifts in form and functions create needs for new kinds of knowledge and skills, we conducted interviews with teachers, eliciting narrative descriptions of how they used these forms and the purposes that they served in their assessment practices. In addition, in these interviews, we also queried teachers about the factors affecting shifts in teachers' uses of assessments.

In the second phase, we fielded a revised survey to a second cohort of K-8 teachers participating in a similar professional development program. Unlike the first cohort, these teachers did not initiate their involvement with their program; these teachers were instead assigned to participate by their schools. Our survey repeated questions on frequency of assessment use, and added new questions about presses adapted from the interview used with our first cohort. These additional items enabled us to sample a greater number of teachers on issues of press.

The two cohorts provided us the opportunity to identify and corroborate general patterns of change in the assessment practices of mathematics teachers who are becoming engaged with reform. Comparisons of the cohorts allowed us to collect preliminary data on both general patterns of change as well as the ways that differences in teachers' reasons for enrollment in professional development programs (initiated vs. assigned) may be related to teachers' experiences of press and to different patterns of change in uses of assessments.

Our study addressed the following questions:

1. How frequently were mathematics teachers utilizing two contrasting forms of assessment tasks (open-ended problems and exercises) and one form of evaluation (rubrics)? Our focus on these three "technologies" enabled us to explore developmental tensions between traditional and

reform-minded assessment methods. While both exercises and open-ended problems are means of eliciting performances from students, the former is typically linked with traditional assessment approaches and the latter with approaches associated with reform. Rubrics are means of evaluating complex performances, and are typically associated with reform.

2. What were the patterns of change in assessment use from last year to this year, and projected from this year to next year?
3. What institutional, stakeholder, and personal factors were affecting shifts in teachers' uses of these assessments?
4. In what ways were the functions of particular forms of assessment changing over time?

Method

Participants

Our first cohort of 35 teachers was engaged in a voluntary 2-year professional development program offered by the UCLA Mathematics project; we administered our survey in the fifth month of the program. They taught kindergarten through twelve grade: Three teachers taught lower elementary, 11 upper elementary, 11 middle school and 10 taught high school.[1] The second cohort of 24 teachers was engaged in a professional development program designed to support their district's system-wide initiative to improve mathematics education. We administered our survey during their initial summer institute. These teachers either volunteered in pairs, or agreed to participate at the request of their principals; they all understood that school participation was required. The teachers taught kindergarten through sixth grade: 10 teachers taught lower elementary, 13 taught upper elementary, and 1 taught middle school.

Measures and Procedures

We developed and administered three instruments: (1) a *survey* to all first cohort teachers, (2) a follow-up *interview* to a subset of 12 of these teachers (six elementary and six secondary), (3) an *integrated survey* for the second cohort that combined items from the prior survey and interview.

---

[1] Lower elementary includes kindergarten through second grade, upper elementary third through fifth grades, middle school sixth though ninth grades and high school tenth through twelve grades.

Survey (for first cohort). The survey requested information on teachers' experience with reform, their interest in implementing reform practices, and the frequency with which they utilized a wide range of methods of assessment. We asked teachers to rate their current use, use last year, and projected use for next year on an eight point Likert scale ranging from 'never use' to 'use daily' (0=never, 1=once or twice per year, 2=three or four times per year, 3=once per month, 4=once or twice per month, 5=once per week, 6=twice or three times per week, 7=daily.) The findings reported in this paper are derived from a subset of the items included on the full surveys, items that pertain to use of exercises, open-ended problems, and rubrics. Appendix A contains key items.

Interview (for subset of first cohort). The interview was partitioned into three parallel sections, one for open-ended problems, another for rubrics, and the final for exercises. In each section, the interview questions were designed to probe teachers' purposes for using a form of assessment, their rationale for shifts in frequency of use, and their perceptions of the factors that affected shifts (or stability) in frequency of use. Thus we asked the teachers to describe how they used each assessment form, what they learned from using it, and how their uses had changed from last year to this year. We then presented teachers with a list of eight factors; we asked them to select one or more of eight possible factors that most influenced any change (or stability) in their use from last year to current practice, rank the selected factors, and explain their rankings. These eight factors included potential "levers for change" (curriculum materials, professional development programs, and district testing), "stakeholder groups" (parents, students, other teachers, and administrators), and "other." The most common reason given for citing "other" was the teachers' own interest—in one teachers words, "my own blossoming thinking!" The protocol for the interview is contained in Appendix B. Interviews were conducted on the telephone by one of two trained project staff members. Interviews required 45-60 minutes.

Integrated survey (for second cohort). The integrated survey used with the second teacher cohort is contained in Appendix C. The items were identical to the initial survey, with the following modifications. First, the items on frequency of use were focused just on exercises, open-ended problems, and rubrics. Second, we included items adapted from the interview; teachers ranked which if any factors (e.g., district testing, administrators, etc.) influenced their use of exercises, open-ended problems, and rubrics.

Results

Our results are organized in three sections. First, we report data on each cohorts' ratings of their engagement with reforms in mathematics education, ratings that are quite high. Second, we report findings on teachers' uses of assessment forms, focusing on current use, trajectories of change, and presses for change/stability. Finally, we present narrative analyses of interviews; the narratives allow for a coordinated examination of the ways that teachers utilize 'old' assessment forms for new purposes or utilize 'new' assessment purposes for familiar purposes, as well as the ways that presses on teachers may impact the forms and functions of their methods of assessment.

Mathematics Teachers' Investment in Reforms

Analyses of teachers' responses to questions about their engagement in reform identified both cohorts of teachers as seriously engaged with reform efforts in mathematics education. Indeed, 94% of the first and 87% of the second teacher cohorts reported a desire to implement the state mathematics frameworks extensively. Further, 66% of the first and 52% of the second teacher cohort characterized their current implementation of the framework as extensive or close to extensive, while another 29% of the first and 44% of the second characterized their implementation as moderate.

Teachers' Use of Assessment Forms: Current Use and Changing Use

Current use. To determine whether there was differential use of *assessment forms* (exercises, open-ended problems, rubrics) in current practice, and whether this pattern varied across our *cohorts* (first cohort [elementary]), second cohort [elementary], and first cohort (secondary), we conducted a 3 (COHORT) x 3 (FORM) ANOVA on teachers' 8-point Likert ratings. The ANOVA revealed a main effect for assessment FORM ($F(2,102)=32.07$, $p<.0001$). Follow-up matched t-tests for the main effect for FORM revealed that teachers reported more frequent uses of exercises than both open-ended problems ($t(df=57)=3.26$, $p<.002$) and rubrics ($t(df=54)=8.45$, $p<.000$), and that more frequent use of open-ended problems than rubrics ($t(df=54)=4.87$, $p<.000$). The effect for GROUP only approached significance ($p<.1$), and there was no FORM x GROUP interaction.

Figure 1 contains a boxplot of teacher frequency ratings for current use of assessment forms. To create the boxplots, we pooled frequency ratings for cohorts, since we found no GROUP or GROUP x FORM interaction effects. The boxplots contain information on the median, quartiles, and extreme gain score values for each group. The "boxes" represent 50% of teachers' ratings that lie between the 25th and 75th percentiles. The boxes' "whiskers" (lines projected from the upper and lower edge of the box) show the high and low scores for the group, excluding moderate and extreme outliers. Moderate outliers (those classrooms with scores between 1.5 and 3 box-lengths from the upper and lower edge of the box) are indicated with an "O," and extreme outliers (classrooms with scores of more than 3 box-lengths from the edges) are indicated with an "X."

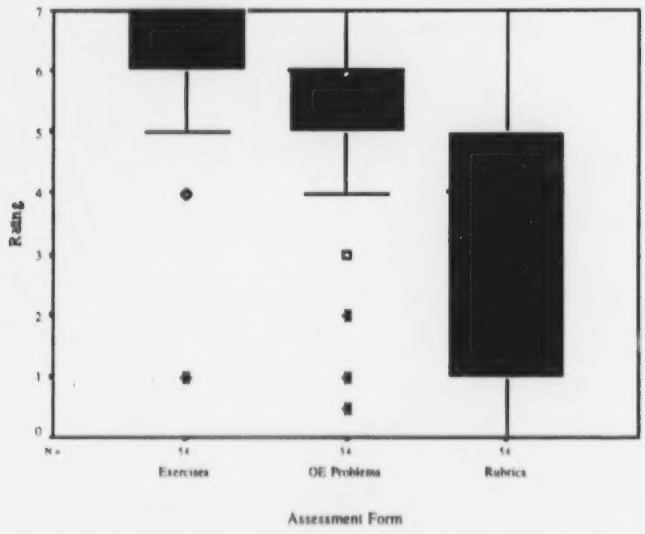

Figure 1. Box Plot of Teacher Frequency Rankings for Current Use of Exercises, Open-ended Problems, and Rubrics.

Figure 1 shows that virtually all teachers in our survey sample reported using exercises frequently for purposes of assessment. Indeed, 75% of the teachers reported using exercises at least 2-3 times a week for assessment. The same was not true for open-ended problems and rubrics: Teachers reported using open-ended problems at more moderate levels, the majority reporting at least weekly use. The variability in use of rubrics was quite pronounced. Indeed, 50% of the

sample reported uses of rubrics in the range between rare (once or twice a year) and relatively frequently (weekly).

When we compared teachers' reported uses of each assessment form in the past, currently, and anticipated in the future, we found that the reported patterns of change were different for each assessment form, as we discuss next.

Change in use. By comparing teachers' reported uses of assessment forms last year, this year, and next year, we were able to identify patterns of change. In our analysis, we coded shifts in frequency from last year to current practice as 'up' if frequency increased, 'stable' if frequency was unchanged, and 'down' if frequency of use declined; we produced a similar coding for shifts in frequency from current to projected practice. These codings produced nine possible trajectories from last year through projected practice. We reduced these nine trajectories into four types: (1) *Increase*—Up-Up, Stable-Up, Up-Stable; (2) *Decrease*—Down-Down, Stable-Down, and Down-Stable; (3) *Stable*—Stable-Stable; and (4) *Mixed*—Up-Down and Down-Up.

For each assessment form, patterns of change were similar for the two cohorts (no chi-squares revealed differences). We therefore pooled cohorts in the bar chart contained in Figure 2. The chart contains the proportion of teachers who showed UP, DOWN, STABLE, or MIXED trajectories for each assessment form.

For exercises, most teachers reported little change in frequency of use. Most already used exercises on a regular basis, and their trajectories show little evidence of decline. Indeed, more than 75% of the teachers reported stable (and high) use over past through prospective practice. In contrast to the results for exercises, most teachers were classified in the UP category for open-ended problems and rubrics. Between 60% and 70% of the teachers' profiles fit an UP trajectory.

Evidence of presses influencing current use. We asked teachers to rank both policy lever factors and stakeholder groups that they felt influenced their current use of exercises, open-ended problems, and rubrics. These data represent the rankings produced by the 12 Cohort 1 teachers that we interviewed, and all of the Cohort 2 teachers. The numerical rankings were supplemented by opportunities for oral (Cohort 1) or written (Cohort 2) commentary on the factors ranked.

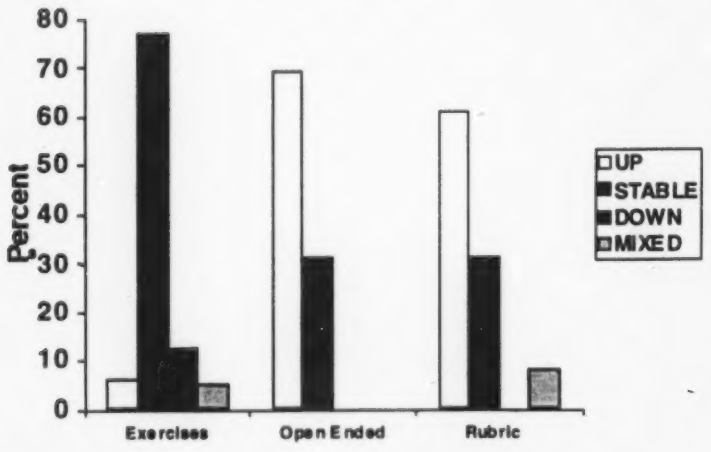

Figure 2. Teacher Trajectories in Frequency of Use of Exercises, Open-ended Problems, and Rubrics.

Because many of the Cohort 1 teachers that we interviewed reported that they found ranking difficult, we ignored the ordinal rankings and treated any ranked categories as reported factors influencing use of the assessment forms. We pooled the results from our two cohorts to increase the size of our sample. Figures 3 and 4 contain bar charts that show the proportion of teachers who ranked a particular lever (Figure 3) or stakeholder group (Figure 4) as a factor influencing their use of exercises, open-ended problems, and rubrics. The results demonstrate that the institutional and stakeholder factors that we listed in our interviews and surveys were indeed perceived by teachers as presses on their assessment practices. However, these factors were perceived by teachers to operate differently across assessment forms. For example, some teachers who cited professional development as a factor indicated that the program in which they were participating advocated a "balanced" approach between exercise-like and more open-ended activities. Of those teachers who cited 'other teachers,'

some indicated that their school colleagues used skills-based approaches while others used inquiry-based approaches.

Figures 3 and 4 show that the teachers interviewed were likely to cite two "levers for change"—curriculum materials and district testing—and two stakeholder groups—students and parents—as factors influencing their decisions to maintain high use of exercises for assessment. *Levers for change* (Figure 3): In their oral and written comments, those teachers who cited curricular materials typically indicated that their texts contained exercises, and those that cited district testing often noted that the tests were often "skills-based" consisting of exercise-like problems. *Stakeholder groups* (Figure 4): Those teachers who selected students cited students' needs to practice skills to perform well on high stakes testing; those who selected parents remarked that parents, in one teacher's words, "want kids to learn the math that they learned."

For use of *open-ended problems* for assessment, teachers were more likely to cite two "levers for change"—curriculum materials and professional development—and two stakeholder groups—students and other teachers. (Recall that teachers' trajectories were variable, though their reports of past, current, and anticipated use of open-ended problems indicated increases in use over time.) *Levers for change* (Figure 3): In their comments, those teachers who cited curriculum materials usually indicated that new texts, replacement units, or materials acquired from professional support groups contained open-ended problems; teachers who cited professional development indicated that these programs had encouraged use of open-ended problems. Of the four teachers who cited district testing, two indicated that their school district had developed a new test that contained open-ended problems. *Stakeholder groups* (Figure 4): Those teachers who cited students typically indicated either that their students preferred open-ended problems for assessment or that their students' knowledge of mathematics grew from using open-ended problems for assessment; those who cited other teachers typically indicated that they had been influenced in talking with teachers who have had success with this form of assessment.

For use of *rubrics* for assessment, teachers were more likely to cite one "lever for change"—professional development—and one stakeholder group—students. *Levers for change* (Figure 3): Those teachers who cited professional development were likely to mention the way that a particular program had supported use of rubrics to evaluate students' responses to open-ended problems.

Stakeholder groups (Figure 4): Those teachers who cited students often explained that use of rubrics makes students' understanding of evaluation "less

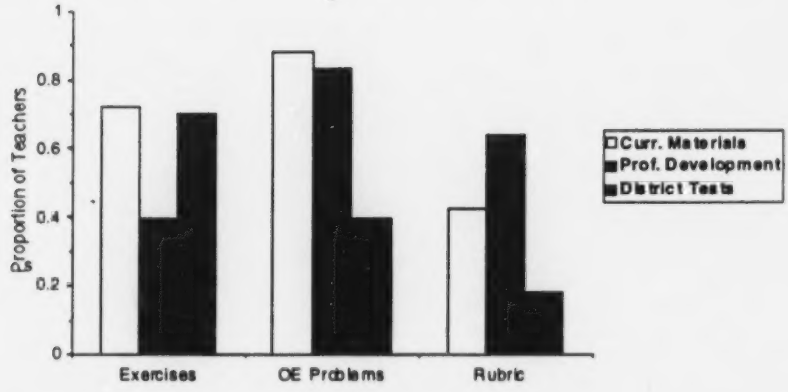

Figure 3. Proportion of teachers citing different "Levers of Change" as presses influencing current users of assessment forms.

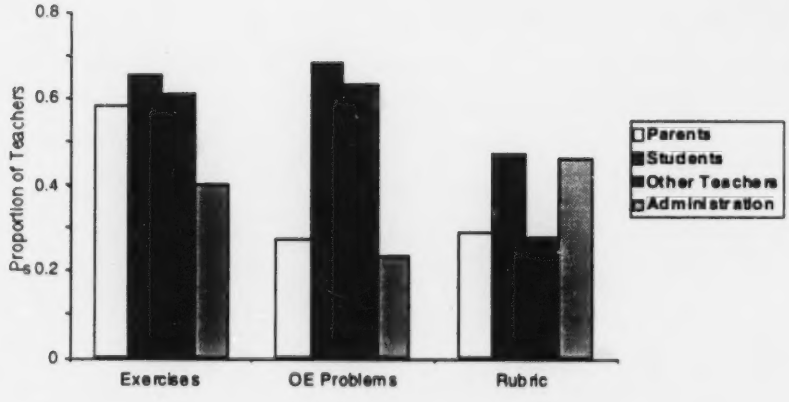

Figure 4. Proportion of teachers citing different stakeholder groups as presses influencing current uses of assessment forms.

of a guessing game." (Similarly, one of the two who cited parents felt that rubrics provided a basis for them to make "subjective grading more concrete to parents.")

Relations Over Time between Assessment Forms and Functions: Two Cases

So far we have considered only shifts in frequency of the assessment forms and the presses that influence frequency of use of these forms. We have not yet considered the assessment functions that teachers were deploying these forms to serve, nor the interplay between the use of particular assessment forms and functions they serve over time.

Our interviews were designed to explore both continuities and discontinuities in forms and functions of assessment. In assessment practices, continuity would be manifested in a teacher's decision to continue using either an 'old' assessment form over time, or, a new form to serve an 'old' function. Discontinuity would be manifested in a teacher's decision to use a new assessment form, or, to use an 'old' form for a new function. Core to our approach is the assumption that continuity and discontinuity are inherently related to one another in the process of development—continuity preserves the coherence or integrity of practice while discontinuity allows for adjustment to presses and organizational change.

To explore the functions of assessment forms for teachers and possible shifting relations between assessment forms and their functions, we analyze two case studies drawn from our interview sample of twelve. The two cases present similarities and contrasts in patterns of change. Though one is an elementary and the other a high school teacher, both illustrate well the interplay between form and function over time in teachers' practices as these teachers work to maintain the coherence of their practice in the context of institutional and stakeholder presses.

Ms. Jones, elementary teacher. Ms. Jones taught a Grade 2–3 split classroom. Throughout her interview, she communicated her interest in change and professional growth—"I'm always looking for new ways of doing assessment, and teaching in general . . ."

Exercises: Repurposing a traditional form to encompass reform functions. The case of Ms. Jones represented continuity in use of an assessment form—

exercises—and discontinuity in function—a shift from a focus on skills and right answers toward a focus on children's understandings of the rationale for skills. She explained that her interest in reform had supported expansion of the functions of assessment in her classroom: "I'm really getting away from the main, old way of doing it. Through that UCLA math program, too, it really explained to me the need for understanding [students' mathematical thinking]." Thus she was beginning to utilize assessment for analysis of student thinking and for instructional planning, but she used tried-and-true exercises as the context for eliciting evidence.

Five or six computation exercises were the focus of Ms. Jones' "morning math activities." Ms. Jones sometimes had students correct their own exercises without making erasures, "so...they show me exactly what it is that they had problems with, and then they get individualized instruction with that difficulty."[2] When probed about what she looked for in a sheet of exercises, Ms. Jones explained that she examined the procedures children used. She offered the example of 21-7=?: If a child were to write down "16," she would know how he produced the calculation—by subtracting seven minus one, instead of one minus seven. Thus, with the support of well-structured exercises, Ms. Jones analyzed students' methods and not just right and wrong solutions. When she then stated that she might use manipulatives to supplement her instruction if a student could not solve the exercises as she intended[3], she demonstrated that she sometimes used her analysis of students' responses as a basis for planning instruction that addressed students' conceptual understandings as well their procedural skills. For Ms. Jones, exercises allowed her to "see how the kids are doing . . . [they give] me a graph on how the child is developing individually." Exercises served a formative function—"it's a tool for myself . . . if I am meeting my objectives, the children are learning, too . . . because then I see how the kids are doing. It allows me to see if I taught it correctly or not."

When comparing her current practice with last year, Ms. Jones reported no change in frequency but changing functions for use of exercises: "I still do my morning math, and I still do my activities; they're done just a little different with

---

[2] Ms. Jones enlisted the help of an aide or a parent to work with individual students.

[3] She referred to the common practice of representing the 'real quantity' of 21 with base-10 blocks, and working through how to 'take away' 7 through an equivalence trade of one 10s block for ten 1s blocks.

the problems." Last year's exercises were tests of knowledge comprehension (retention of taught skills), while this year's enabled her to assess "higher order thinking." She attributed this shift in function to the UCLA professional development program that had focused her on "problem-solving, logical reasoning—I think now my [classroom] program is more geared to develop those in students than it was probably before."

Ms. Jones did not anticipate changing her use of exercises for assessment purposes next year. Pleased with the new ways she was using exercises to assess 'higher order thinking,' she saw no reason to change.

Open-ended problems: Opportunities for discovery. Ms. Jones had been encouraged to use open-ended problems in her professional development program, and she found many open-ended problems in the new curriculum materials her school had adopted. Her "own changing views and blossoming through my professional development" contributed to her growing interest in incorporating open-ended problems, a new form of practice, into her instructional program. Thus Ms. Jones expressed delight at her students' mathematical discoveries and the potential of open-ended problems for student learning.

> I use a lot more [open-ended problems] than I did last year . . . and I'm really seeing there is a change in the students by doing so much. I see them coming up with things and noticing patterns. Things that I really don't notice, they find, and to me that's amazing. . . . I think it's because I'm letting them think more. Instead of having a direct answer that is grading for the answer, I think the kids are having to see more, and I think they're blooming with the opportunity to do that.

She focused on the pleasure she and her students derived from the diversity of strategies students constructed when solving these kinds of problems.

She was not relying much on open-ended problems as a context for eliciting and evaluating students' mathematical understandings and skills. When she described one effort to use open-ended problems for assessment, her description suggested that she was using this new form for a prior assessment function—she was evaluating whether students' answers were right or wrong, just as she used to do with students' exercise sheets. In the example below, she explained how she used an estimation jar activity to determine which students had no understanding at all of estimation:

> Children not only give their guess but they have to explain to me the reasoning on their guess. And they have to write it out, the process that they use, and then they sometimes do an illustration of it. . . . I put (the estimates) on a big bulletin board, and then they glue it onto this section to see how close children are for the right answer, but I can also see where the children are completely off.

Last year Ms. Jones used open-ended problems less frequently, and she rarely if ever used them as an opportunity to analyze student thinking: "I might have looked at them, but I don't think I looked at them as deeply." She planned next year to implement a new form of mathematics assessment task—long-term investigations—but she did not report that she planned to use investigations to elicit and analyze student understanding or skill:

> I'd like to be a little more daring. Instead of doing all open-ended things, like every day, like I do (now), I'd like to take one large project and expand on it and allow the children to have that expansion time. Or at least go a month. . . . Because we do things now . . . where we're doing measurement, and we do hands-on a lot, and we do a lot of open-ended questions. . . . I think I'd like to take them through the whole carry-through, . . .

Thus it appeared that there would be continuity in the function of her open-ended tasks—the instructional function of encouraging discovery The shift would be one in form (addition of investigations to her program), not in function.

Rubrics: Focus on the quality of explanation. Ms. Jones had tried using a rubric for the first time this year: "They [rubrics] scared me. It was new, and I'd never done it before." Interested in working with her colleagues, she started with one rubric designed by teachers at her school and supported by her principal: "Yes, I've looked at [other rubrics], but right now I'm just trying to get a grasp on using [this] rubric." She had been encouraged by the staff of her professional development program as well as the representative from her school's new textbook series who modeled using rubrics for assessment.

Ms. Jones felt that her colleagues' rubric provided a framework for evaluating students' responses to open-ended problems, a framework that she felt was missing in the comments she used to give. The rubric had four levels. While a criterion for each level included a global judgment of students' understanding of the task, there was particular importance placed on the quality of the explanation—inclusion of detail and examples.

> A star is the highest, a happy face, a check, and a minus. . . . If I ask the question of multiplication, 'what is multiplication,' if the child is completely off his or her rocker and writes nothing, that would be my minus, obviously, because then they don't have any of the concept to grasp. If the child can answer the question about 'what is multiplication' by, you know, 'it's a way of grouping things,' that would be considered a check. If a child writes 'it's a way of grouping items—for example if I have two baskets and each basket has three oranges in it, it would equal six' . . . if the child has not only given me a definition but has added a little bit more to the definition . . . with the explanation, then they get a happy face. . . . And then my star would be someone who is really clear and precise, has the definition but also has say, for example, two or more examples, so I'm able to see that the whole understanding process is there.

Ms. Jones felt that students who received a star or a happy face both had understanding; these levels of performance were distinguished by the amount of explanation detail.

> Well, it's hard to explain, because once you see the differences in the papers you see the differences in the papers. I want to call it more juicy, that my star is really, really juicy, with a lot of information and a lot of detail, and I can see a really well thought-out process.

Intent on learning to use this rubric as it was, Ms. Jones was not concerned with its weakness as a support for evaluating mathematical thinking. Indeed, she linked rubrics to her prior reliance on "percent correct" when she said, "In a way [the rubric is] sort of based on percentage, because they have to show me certain skills to qualify for their number that they receive on their rubric." Ms. Jones was committed to continued use of rubrics. While the impetus for implementing a rubric was influenced by individuals outside her classroom (colleagues, principal, professional developers, textbook representative), her commitment reflected her perceptions of the usefulness of rubrics within her classroom. First, she had come to believe that a score such as percent correct was not appropriate for evaluating open-ended problems: "For me, personally, [rubrics are] probably one of the only ways to grade [students' responses to open-ended problems], because [such responses are] so varied." Second, she had observed how useful rubrics had been in communications with her students and their parents. She felt that her students worked harder when they knew how their open-ended problems were evaluated, and that parents had a better understanding "why this child got the grade he or she did." She explained to the parents, "Well, this is

what I'm looking for here, and as you can see here your child is showing me this."

Enthused about rubrics, next year Ms. Jones expected to continue to use rubrics but anticipated shifting in how she used them as she gained competence and facility with scoring. The shift that she anticipated was a shift in efficiency or skill, not a shift in assessment function.

> Hopefully I'd get better at doing it. Then I'd be using them more often, because right now not every single paper that I receive is graded by a rubric. It will be checked off if the child does it, but . . . it takes a lot of time for me right now, still, to sit down and do it.

Ms. Jones hoped that next year she could better manage the time entailed in scoring, but she otherwise planned to use the same rubric to capture the same aspects of students' work.

Ms. Smith: High school teacher. As in the case of Ms. Jones, Ms. Smith's uses of assessment forms and the assessment functions that they served played off one another over the course of her evolving practice. Ms. Smith, like Ms. Jones, made an effort to assess students' understandings of the rationale for procedures, in part by asking students to explain their procedures in writing. When it came to open-ended problems and rubrics, however, Ms. Smith, illustrated a different trajectory, one in which these new forms of assessment were beginning to serve the function of eliciting and analyzing students' mathematical thinking.

Exercises: A focus on misconceptions as well as accuracy. Like many of the teachers in our sample, Ms. Smith cited curriculum materials, district testing, students, and parents as presses that influenced the frequency of her use of exercises for assessment purposes. In her new curriculum materials, there were more "hands-on activities," but "then they do some exercises relating to those activities." Wanting her students to do well on high stakes assessments, Ms. Smith explained that "district testing [that] has multiple choice problems, which are more of these exercise type problems," and thus her students "need the exercises in order to practice. . . and to feel more comfortable with [the test]." She commented as well that some parents think mathematics is like basic exercises, "So I guess they have to see some of those or they wouldn't think it's any mathematics."

Thus Ms. Jones used exercises frequently for assessment, but she also noted that she found it difficult to use exercises to gain insight into student thinking: "It's hard to see with an exercise anything else [other than accuracy] . . . " Her assessment strategy was to determine whether students' answers were correct; if they were not correct, she tried to determine whether "there's a misconception of something." To help her identify misconceptions, this year, compared with last year, she had begun asking students "to explain . . . like, 'Problem number five, how did you do it?' So that I get a feel of what they're doing. And for me it was to get them to do more writing and to understand their thought process." Thus, like Ms. Jones, she had supplemented exercises with written explanations of procedures to help her identify student thinking.

For next year, Ms. Smith did not anticipate shifting either the frequency of exercises or her methods of evaluating students' misconceptions of exercise procedures. She was pleased with her current use of exercises for assessment.

Open-ended problems: Focus on strategy and projected focus on domain. Ms. Smith reported that her interest in using open-ended problems to elicit evidence of students' mathematical thinking had grown over the last year. Last year she just started with a new curriculum, "so I pretty much kind of followed what I needed to do first. And this year, since I'm used to [the] curriculum, I'm doing more things on my own . . . much more [student] writing . . . I can see more of their thought processes." She was also concerned to prepare her students for her district's annual performance-based tests.

> Part of the testing has open-ended questions. So I don't feel preparing them a day ahead or two days ahead, which is (what we're supposed to do), will prepare any student for any kind of writing if they haven't been doing it in class already. So I made it a point to have them do more writing, to make them more comfortable when they take tests . . . so it's more second nature than 'oh, my gosh, here's a math problem and I have to solve it by writing and I've never done it.'

Thus she established writing as an important mode of expression in mathematics, and used writing as evidence of students' mathematical thinking.

> It's easier to grade [students' responses to open-ended problems] and it's easier to look at it when I'm looking at how thoroughly they understand it in their thinking process. And I can get a better idea when they write it in words than if they just write it in numbers. Because my question sometimes is, 'Where are they getting these numbers from, if I don't know their understanding of it?' So by them writing down and

> thoroughly writing their thoughts down I can easily see where the misconceptions are, if there are any, or I can see where they're taking the problem.

Ms. Smith regarded the two most important goals for evaluating students' responses to ended problems as "strategies" and "communicating what they understand."

Ms. Smith's plans for next year suggested continuities in her use of the open-ended task form for the function of eliciting students' mathematical thinking. Indeed, she planned to expand her use of open-ended problems by assigning students a series of problems over time to track progress in skills and understandings in specific mathematical domains. Below she outlines her plan to gather evidence of students' progress in understanding functions.

> Usually on a traditional test, or just in assessment in general, you're assessing maybe things that you've covered. So what I would do is . . . for example, like if we're doing . . . distributive property and graphing, I might ask more of an open-ended problem to see how they've progressed. And what I would want to do with this specific class is to [use a] growth problem where I give them the same problem over a period of time, and see how much they progress. . . . [The problem] would be, like, 'Tell me all you know about a function.' So in the beginning I would give them two functions, they would have to graph it. Minimally they'll be just graphing it and maybe doing a table or something. And as we progress on with the class they might be putting like the domain and range into the problem, talking about symmetry, axis of symmetry . . . move on that way. And as they get more sophisticated in what they know about the problem, they would be adding more to the problem. So from the beginning to the end of the year, they could see how much they've progressed in terms of the mathematics.

Ms. Smith was planning to continue utilizing the 'tools' of her current practice to design a more comprehensive set of assessments more deeply grounded in the mathematics of her courses. Ms. Smith was expanding her conception of assessable domains, designing methods to assess student progress within each domain, and planning to use a variety of assessment forms (each of which were already in her assessment 'repertoire') to capture different kinds of knowledge and skill within each domain.

Rubrics: Focus on problem solving, tension between richness of rubric content and efficiency in scoring. Ms. Smith used a rubric in her current practice to evaluate her students' work on the "problem of the week." She explained that

she developed the rubric the prior year; she appropriated a rubric from a colleague and redesigned it to suit her needs.

> Well, actually, I stole it from someone, so I didn't design it myself. But it started off where the person I borrowed it from had a fifteen-point rubric, and I didn't feel that it went in line with what I wanted them to do so I kind of adjusted it. . . . I've taken that person's fifteen points, and then some points from different workshops, you know, other people's rubrics and the CLAS rubric that they used to have . . . that kind of put it together for something that I felt comfortable with and I felt that the students could look at and use.

She was motivated to use the rubric to help students "[see] their understandings"; she compared the rubric to district testing, arguing that it "kind of forces students to see what's expected of them, or what they should know." She also cited her interest in becoming more engaged with the mathematics education community, "just to kind of align myself more with what's going on with mathematics, and to get out of the tradition of just testing and looking for numbers and looking for right answers, and more looking for the process. . . . it's not just the answer that's important, but the processes."

Ms. Smith used the rubric as a mechanism for setting a standard, "letting students know what they need to do [on the problem of the week] to achieve a grade. More so than just saying, well, you know, you get five points for this if it's correct . . . It's more like, well, I'm looking for a [quality] type of thing." Students "have a whole week . . . so they have time to kind of look at [the rubric], and throughout the week I kind of have them look back on it." Ms. Smith designed the rubric to convey "what I want from them." Her 10-point scale consisted of four components that encompassed stages of problem solving and analysis: (a) restating the problem (2 points) , (b) strategy (4 points), (c) solution (2 points), and (d) reflection (2 points). In the four-point subscheme for strategy, for example, "zero would be 'you didn't show anything,' one would be just maybe putting down a few numbers, no attempt to really solve the problem. And it would go all the way to four, which would be a complete solution and asking what the whole problem asks for. Sometimes in my problem I'll say, 'Give me two solutions,' or I might say, 'you have to draw.'"

Ms. Smith anticipated using rubrics next year, though she was considering some modification. While maintaining her interest in student thinking, she was

considering adopting or adapting the rubric provided with the curriculum she was implementing.

> In the . . . program, they grade all types of problems, like test problems, on a four-point scale, and I've seen people use that, and I'm not yet comfortable with it. So it might be something I try. I think it's a more holistic . . . a little easier to grade. . . . Teachers who've been using it say it's a little bit easier for them to grade than to have to, like, nit-pick things.

Ms. Smith was concerned with developing a method of rubric use that was less time-consuming; less time per problem could mean more time to score more types of problems than just the problem of the week. Ms. Smith recognized that the use of the simpler form meant sacrificing some of the components of her current rubric and producing less information about student thinking. As she pondered her plans for the coming year, Ms. Smith was struggling with the trade-offs for teachers between qualitative analysis and expediency.

Patterns of change in the two cases. These two cases illustrate different patterns of development. Both Ms. Jones and Ms. Smith re-purposed their uses of exercises to allow them to assess students' procedures as well as students' understandings of the procedures—they examined patterns of responses to sets of exercises as well as students' written explanations of their procedures. The teachers differed, however, in their changing uses of open-ended problems and rubrics. Ms. Jones viewed open-ended tasks principally in terms of instructional functions, as opportunities for student discovery; when she had time to evaluate students' responses with a rubric, she focused on the correctness of the solution or on the quality of the written explanations more than the quality of students' mathematical understandings. In contrast, Ms. Smith viewed open-ended tasks as opportunities to gain insight into her students' misconceptions; she assigned these tasks once a week, and evaluated the responses with a rubric designed to capture students' competence with phases of problem solving.

Using these two cases, we documented several patterns of development. None of the patterns represent a radical re-organization of practice. Rather, for each pattern, development is marked by both continuity and discontinuity.

One pattern captures the ways that teachers may implement *a new form of assessment in a way that served 'old' functions.* Ms. Jones used a 'new' form of assessment, open-ended problems, in ways that served instructional function.

She engaged children with the open-ended problems to provide them the opportunity to invent strategies; she did not examine students' responses to open-ended problems to gain insight into the character of their mathematical thinking, a function linked to student inquiry promoted by reform documents.

A second pattern captures the ways that *new forms of assessment may be implemented in pro forma ways.* Ms. Jones used a rubric developed by colleagues—a rubric that focused on the completeness of a student's written explanation. She did not revise it to capture students' mathematics.

A third pattern illustrates the ways that *teachers may fashion or re-fashion forms of assessment in order to assess* students' *mathematical thinking,* the function of assessment recommended by reform. Both Ms. Jones and Ms. Smith re-purposed an 'old' form of assessment, an exercise, to serve a new function, supplementing the old form as necessary with new forms (written explanations) that support the new function. In addition, Ms. Smith appropriated a colleague's rubric for evaluating students' responses to the open-ended problem of the week, and then redesigned it to suit her curriculum and her goals for her students' mathematical learning.

A fourth pattern illustrates how teachers' *concerns for efficiency* may work against the quality of their assessments. Both teachers were considering strategies for more frequent and more rapid rubric scoring. Ms. Jones as yet had no specific strategy for increasing the speed of scoring; Ms. Smith was considering replacing her analytic rubric with a holistic approach, and she expressed worries about tradeoffs between frequency of scoring and quality of the evaluation.

Discussion and Concluding Remarks

Our efforts were guided by a framework for understanding the professional development of teachers who are invested in current educational reforms in mathematics. We assumed that teachers construct their assessment practices on a daily basis, sustaining a network of routines in classroom activities as they adjust to or resist a matrix presses for change. In our study, we collected the self-reports of two cohorts of reform-minded teachers regarding their uses of three assessment forms—exercises, a staple of traditional instruction, and open-ended problems and rubrics, both valued in current reform efforts. We analyzed both

frequency of use and patterns of developmental change in the forms and functions of assessment as teachers were engaged with ongoing presses.

Frequency of Use

The two cohorts of teachers reported similar patterns of frequency of use of each assessment form, and similar patterns of changing use. Exercises, the staple of traditional assessment practices, were used at high frequency levels by most teachers, and there was no anticipation of a decrease in use. Open-ended problems were used at moderate levels of frequency, and use was 'on the rise'; compared with exercises, there was somewhat greater variation among the teachers in current use and in change in use. The findings for rubrics were the most variable, with many teachers reporting fairly low levels of use and with much inconsistency among teachers in projected use. Teachers' reports of the presses on their assessment choices provided some explanation of these frequency patterns. Teachers were likely to cite a substantial number of converging institutional and stakeholder presses to use both exercises (i.e., a press to maintain high use) and open-ended problems (a press to increase use). They cited fewer categories of press to use rubrics, mentioning most often their current off-site professional development program.

The pattern of findings for frequency of use suggests that, while mathematics teachers are increasingly likely to assign open-ended problems to elicit students' mathematical thinking, they are less likely to evaluate students' responses to those problems with rubrics. This infrequent use of rubrics appears to reflect less press to use them. It is a worrisome finding. While rubrics are not the only means of evaluating complex student performance, they are an important strategy for representing the content and quality of students' mathematical thinking and learning. If our findings suggest that teachers are eliciting but not evaluating students' responses to complex problems, then teachers are missing critical opportunities for building instruction on evidence of student learning.

Patterns of Developmental Change

Our case analyses provided evidence of the pathways by which teachers implement new forms of assessment, or develop new functions for existing methods of assessment.

On the one hand, teachers may use an 'old form' of assessment for a 'new function.' Both of our case teachers reported building instruction on an analysis of their students' understandings of exercises, an 'old' form of assessment; they were no longer limiting their analysis of student learning to the percentage of correct answers. This finding has implications for classroom practice as well as strategies for building teacher capacity. Exercises are well-constrained tasks with which teachers and students are very familiar; teachers have developed considerable understanding of the conceptual hurdles that confront children as they engage with exercises and work to gain understanding and skill. Encouraging teachers and students to examine the thinking that underlies students' responses to exercises represents one pathway to improvement in assessment practices.

On the other hand, teachers may implement a 'new form' of assessment to serve an 'old function.' We found that some teachers posed open-ended problems, a new kind of task, and then evaluated the responses as correct vs. incorrect, an 'old' method of scoring student work. The implication of this pattern is that teachers may benefit from opportunities to consider the ways that new forms of assessment afford them insights into students' mathematical understandings.

Our case analyses suggest that the contents and forms of assessments constrain the kinds of insights teachers are likely to construct. When teachers implement rubric scoring, for example, the scores they produce may not represent an analysis of students' mathematical thinking that is an adequate basis on which to build further mathematics instruction. A rubric that represents substantive aspects of children's mathematics is more likely to provide a frame to guide teachers' interpretations. Such rubrics are also more likely to prompt a teacher to reconstruct his or her goals and methods of assessment. The rubric that Ms. Jones adopted, for example, a rubric that focused on quality of writing, did not challenge her to reconstruct her goals. Thus, 'learning how to use a rubric' represented a discontinuity in form (adoption of rubric scoring) and a continuity in function (celebrating discoveries or assessing countable skills). We believe that the pathway of her development would have been different if the rubric had pressed for greater analysis of children's mathematical thinking.

The burden of scoring student work with rubrics may become a press to use them less, or, to use a simpler rubric. Ms. Smith was considering replacing her

analytic rubric for students' work on 'problem of the week' with a holistic one. We worry that her goal to increase the frequency of rubric scoring, using a simpler rubric, will result in a shallower analysis of her students' mathematical thinking. Ms. Smith's dilemma makes clear that the capacity of assessment to support sound instruction depends on the *feasibility* of the methods. When we consider developmental relations over time in teachers' uses of particular methods of assessment, we must include consideration of the ways that teachers' goals reflect the constraints of large class sizes and heavy teaching loads.

Research on teacher cognition and the implementation of new practices often concludes with the maxim that "change takes time." In order to understand why 'change takes time,' we need to identify developmental patterns in the ways that teachers construct goals for their practices, goals that interweave the presses upon them, the resources available to them, and their current knowledge and patterns of practice. Our study demonstrates the importance of examining the dynamics of change in the professional development of teachers.

References

California State Department of Education. (1992). *Mathematics Framework for California Public Schools: K-12.* Sacramento, CA.

Fennema, E., Carpenter, T. P., Franke, M. L., & Carey, D. A. (1992). Learning to use children's mathematics thinking. A case study. In R. Davis and C. Maher (Eds.), *Schools, Mathematics, and the World of Reality* (pp. 93-117). Needham Heights, MA: Allyn & Bacon.

Goldsmith, L, & Schifter, D. (1997). Understanding teachers in transition: Characteristics of a model for developing teachers. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 19-54). Lawrence Erlbaum Associates: Mahwah, New Jersey.

Jones, D. (1997). A conceptual framework for studying the relevance of context to mathematics teachers' change. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 131-154). Lawrence Erlbaum Associates: Mahwah, New Jersey.

National Council of Teachers of Mathematics. (1993). *Professional Teaching Standards.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

Nelson, B. S. (1997). Learning about teacher change in the context of mathematics reform: Where have we come from? In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 3-15). Lawrence Erlbaum Associates: Mahwah, New Jersey.

Scribner, S. & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-22.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 127-146). New York: MacMillan Publishing Company.

Tyack, D. B., & Cuban L. (1995). *Tinkering toward utopia: A century of public school reform.* Cambridge, MA: Harvard University Press.

Wood, T., Cobb, P. & Yackel, E. (1991). Change in teaching mathematics: A case study. *American Educational Research Journal, 28,* 587-616.

APPENDIX A

Appendix A

Survey Items Utilized In This Report

YOUR BACKGROUND

Name _____

Experience:

Number of years teaching mathematics at any grade level _____

Please rank 1-5.	None		Some		Extensive
How would you characterize your implementation of the California State Framework in your classroom?	1	2	3	4	5
How would you characterize your desire to implement the California State Framework in your classroom?	1	2	3	4	5

Grade level(s) / courses you teach this year:

_____

_____

_____

YOUR METHODS OF ASSESSMENT

Assessment tasks and problems:

Please estimate (a) the frequency with which you use these options for assessment purposes currently, (b) your use last year , and (c) your expected use next year.

	Frequency of use				
A.	**Current**				
-exercises (e.g., computation; short, structured problems)	Daily	2-3/wk	1/wk	1-2/mo	1/mo
	3-4/yr	1-2/yr	never		
-open-ended problems	Daily	2-3/wk	1/wk	1-2/mo	1/mo
	3-4/yr	1-2/yr	never		

B.	**Last Year**				
-exercises (e.g., computation; short, structured problems)	Daily	2-3/wk	1/wk	1-2/mo	1/mo
	3-4/yr	1-2/yr	never		
-open-ended problems	Daily	2-3/wk	1/wk	1-2/mo	1/mo
	3-4/yr	1-2/yr	never		

C.	**Next Year**				
-exercises (e.g., computation; short, structured problems)	Daily	2-3/wk	1/wk	1-2/mo	1/mo
	3-4/yr	1-2/yr	never		
-open-ended problems	Daily	2-3/wk	1/wk	1-2/mo	1/mo
	3-4/yr	1-2/yr	never		

Methods of feedback to students:

Please estimate (a) the frequency with which you use the following methods of feedback currently, (b) your use last year , and (c) your expected use next year.

	Frequency of use
A.	**Current**
-rubric score	Daily 2-3/wk 1/wk 1-2/mo 1/mo 3-4/yr 1-2/yr never

B.	**Last Year**
-rubric score	Daily 2-3/wk 1/wk 1-2/mo 1/mo 3-4/yr 1-2/yr never

C.	**Next Year**
-rubric score	Daily 2-3/wk 1/wk 1-2/mo 1/mo 3-4/yr 1-2/yr never

APPENDIX B

Appendix B

INTERVIEW: FOLLOW-UP FOR DIFFERENT SURVEY PROFILES

We're interested in assessment in a broad sense. We're interested in the ways teachers assess what students know and can do in math. We know that you may use a variety of ways to assess what your students know. We're going to focus on two types of tasks—exercises and open-ended problems.

I. OPEN-ENDED PROBLEMS: SHIFTS IN FORMS & FUNCTIONS

A. If I were to sit in your classroom over the course of a week, what would I see in terms of how you use open-ended problems for assessment purposes?

1. What do you learn from this?
2. How does that provide you with information about your students?

B. Would I have seen you using open-ended problems differently for assessment purposes last year?

1. How?
2. Why?

C. Would I see you using open-ended problems differently for assessment purposes next year?

1. How?
2. Why?

II. EVALUATING OPEN-ENDED PROBLEMS USING RUBRICS

A. Do you ever evaluate or provide feedback in the form of rubrics to open-ended problems? (If not: Okay, well I'd still like to understand what may have influenced your decision not to use rubrics)

1. Didn't use rubrics because of:

a) ____Curriculum materials

b) ____School administration

c) ____Parents

d) ____District testing

e) ____Professional development program

f) ____Students

g) ____Other teachers

h) ____Other

B. (If rubric used:)

1. What are you looking for when you use a rubric?
2. What do your levels designate?
3. Would I have seen you using rubrics to evaluate open-ended problems last year?
   a) How?
   b) Why?
4. Did you use rubrics more or less frequently last year compared to this year to evaluate open-ended problems?
5. I noticed that last year you used rubrics [more / less / same] *frequently for evaluating open-ended problems.* Please take a look at part II-E of the handout. Did any of the following factors influence your change or stability in frequency of use? If so, please rank them in order of importance. Let 1= the most influence.
   a) Exercises
      (1) ____Curriculum materials
      (2) ____School administration
      (3) ____Parents
      (4) ____District testing
      (5) ____Professional development program
      (6) ____Students
      (7) ____Other teachers
      (8) ____Other
6. Would I see you using rubrics next year?
   a) How?
   b) Why?
   c) Do you expect to use rubrics more or less frequently next year compared to this year to evaluate open-ended problems?

III. Exercises: Shifts in Forms & Functions

A. If I were to sit in your classroom over the course of a week, what would I see in terms of how you use exercises for assessment purposes?

1. What do you learn from this?
2. How does that provide you with information about your students?

B. Would I have seen you using exercises differently for assessment purposes last year?

1. How?
2. Why?

C. Would I see you using exercises differently for assessment purposes next year?

1. How?
2. Why?

IV. Factors Influencing Shifts in Frequency for Open-ended Problems

A. I noticed that last year you used open-ended problems [*more / less / same] frequently* for *assessment purposes*. Please take a look at part I-E of the handout. Did any of the following factors influence your change or stability in frequency of use? If so, please rank them in order of importance. Let 1= the most influence.

B. Open-ended problems

1. ____Curriculum materials
2. ____School administration
3. ____Parents
4. ____District testing
5. ____Professional development program
6. ____Students
7. ____Other teachers
8. ____Other

V. Factors Influencing Shifts in Frequency for Rubrics

A. I noticed that last year you used rubrics [*more / less / same] frequently* for *assessment purposes*. Please take a look at part I-E of the handout. Did any of the following factors influence your change or stability in

frequency of use? If so, please rank them in order of importance. Let 1= the most influence.

B. Rubrics

1. ____Curriculum materials
2. ____School administration
3. ____Parents
4. ____District testing
5. ____Professional development program
6. ____Students
7. ____Other teachers
8. ____Other

VI. Factors Influencing Shifts in Frequency for Exercises

A. I noticed that last year you used exercises [more / less / same] *frequently* for *assessment purposes*. Please take a look at part I-E of the handout. Did any of the following factors influence your change or stability in frequency of use? If so, please rank them in order of importance. Let 1 = the most influence.

1. Exercises
   a) ____Curriculum materials
   b) ____School administration
   c) ____Parents
   d) ____District testing
   e) ____Professional development program
   f) ____Students
   g) ____Other teachers
   h) ____Other

VII. Feedback

A. Do you use exercises or open-ended problems:

1. To get information for providing feedback to parents?
2. Was this any different last year? How?
3. Will it be any different next year? How?

APPENDIX C

APPENDIX C

Teacher Questionnaire:

Math Assessment in Your Classroom

May, 1997

Dear Teachers:

The information you provide us on the attached survey will help us understand how math teachers are assessing their students' learning in the classroom. In the last decade, there have been many changes in classroom assessment, and teachers are facing the challenge of choosing what kinds of assessment methods to use. Your response to this survey will give us useful information on what teachers are choosing to use and the factors that influence their choices.

The information you provide will be confidential and available only to members of the CRESST research team at the University of California, Los Angeles (UCLA). When we publish reports of the research, we will make no mention of the actual names of the schools or specific people who responded to this survey. Your participation is voluntary, and you may choose not to answer questions.

If you have any questions or concerns, please contact Maryl Gearhart at (310) 206-4320 or maryl@cse.ucla.edu.

Thank you very much for taking the time to complete the survey.

CRESST Research Staff:

Megan L. Franke, Assistant Professor, Graduate School of Education

Maryl Gearhart, Project Director, CRESST

Geoffrey B. Saxe, Professor, Graduate School of Education

Stephanie Biagetti, Research Associate

Lisa Butler, Research Associate

Michele Crockett, Research Associate

Sharon Howard, Research Associate

Linda St. John, Post-doctoral Research Associate

MATH ASSESSMENT IN YOUR CLASSROOM

Name _____

School _____

District _____

Grade level(s) and courses you teach this year:

Grade level(s)	*Course(s)*
_____	_____
_____	_____
_____	_____

Curriculum in use:

Please identify the math curriculum you are using this year.

Textbook and teacher's guide:

_____

_____

_____

_____

Replacement units, if any:

Additional resources, if any:

_____

_____

Implementation of Framework this year, last year, next year:

	None		Some		Extensive
How would you characterize your implementation of the California State Framework in your classroom *this year?*	1	2	3	4	5
How would you characterize your implementation of the California State Framework in your classroom *last year?*	1	2	3	4	5
How would you characterize your goals for implementation of the California State Framework in your classroom *next year?*	1	2	3	4	5

Professional development in mathematics education

Please indicate the number of sessions you've attended over the last two years.

	Number of sessions attended over the last 2 years			
Math curriculum training	0	1	2	3 or more
Training in math replacement units	0	1	2	3 or more
Math Project (e.g., UCLA, Dominguez Hills)	0	1	2	3 or more
Participation in LAUSD's LA-SI (math)	0	1	2	3 or more
School workshops and staff development in math education	0	1	2	3 or more
District workshops and staff development in math education	0	1	2	3 or more
County workshops and staff development in math education	0	1	2	3 or more
Off-site professional conferences in math education	0	1	2	3 or more
Other: _____	0	1	2	3 or more

FOR THE REMAINDER OF THIS SURVEY, PICK ONE GRADE LEVEL AND/OR COURSE, AND ANSWER QUESTIONS PERTAINING TO THAT GRADE/COURSE.

WHAT GRADE/COURSE DID YOU PICK? _____

Assessment types:

Exercises, Open-ended problems, Projects/Investigations, Portfolios

Please estimate (a) the frequency with which you use these options for assessment purposes currently,

(b) your use last year, (c) your project used next year.

	Frequency of use: Current				
-exercises (e.g., short, structured problems that assess computation procedures)	Daily	2-3/wk	1/wk	1-2/mo	3-4/yr
	1-2/yr	never			
-open-ended problems (e.g., problems that assess multiple approaches, multiple skills and concepts)	Daily	2-3/wk	1/wk	1-2/mo	3-4/yr
	1-2/yr	never			
-math projects or investigations (e.g., long-term projects that engage students with multiple approaches, skills, concepts, applications)	Daily	2-3/wk	1/wk	1-2/mo	3-4/yr
	1-2/yr	never			
-student math portfolios (e.g., presentation of student work for the purpose of showing achievement)	Daily	2-3/wk	1/wk	1-2/mo	3-4/yr
	1-2/yr	never			

	Frequency of use: Last Year
-exercises (e.g., short, structured problems that assess computation procedures)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never
-open-ended problems (e.g., problems that assess multiple approaches, multiple skills and concepts)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never
-math projects or investigations (e.g., long-term projects that engage students with multiple approaches, skills, concepts, applications)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never
-student math portfolios (e.g., presentation of student work for the purpose of showing achievement)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never

	Frequency of use: Next Year
-exercises (e.g., short, structured problems that assess computation procedures)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never
-open-ended problems (e.g., problems that assess multiple approaches, multiple skills and concepts)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never
-math projects or investigations (e.g., long-term projects that engage students with multiple approaches, skills, concepts, applications)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never
-student math portfolios (e.g., presentation of student work for the purpose of showing achievement)	Daily 2-3/wk 1/wk 1-2/mo 3-4/yr 1-2/yr never

Methods of Feedback to Students:

Please estimate (a) the frequency with which you use these feedback options currently, (b) your use last year, and (c) your projected use next year.

A.	*Frequency of use: Current*					
-score (% or number correct)	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-letter grade	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-rubric score	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-written feedback	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-oral feedback to individual student	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-other:_____	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr

B.	*Frequency of use: Last Year*					
-score (% or number correct)	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-letter grade	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-rubric score	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-written feedback	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-oral feedback to individual student	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-other:_____	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr

	Frequency of use: Next Year					
-score (% or number correct)	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-letter grade	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-rubric score	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-written feedback	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-oral feedback to individual student	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr
-other:_____	Daily never	2-3/wk	1/wk	1-2/mo	3-4/yr	1-2/yr

Factors that influence your methods of assessment: Teachers have told us that some of the factors listed below influence their decisions about math assessment. How critical are any of these factors in *your* current decisions to use a particular approach to math assessment?

Please rank any factor(s) that apply to your current decisions about math assessment. Let 1 = the most influence.

Explain how the factors you ranked influence your current decisions about math assessment.

EXERCISES:	
___ Curriculum materials	
___ District testing	
___ Other teachers	
___ Parents	
___ Professional development	
___ School administration	
___ Students	
___ Other _____	
OPEN-ENDED PROBLEMS:	
___ Curriculum materials	
___ District testing	
___ Other teachers	
___ Parents	

___ Professional development	
___ School administration	
___ Students	
___ Other _____	
RUBRICS:	
___ Curriculum materials	
___ District testing	
___ Other teachers	
___ Parents	
___ Professional development	
___ School administration	
___ Students	
___ Other _____	

THANK YOU!

From: Bev Huff (4/28/98)
To: Kim Hurst

Mail*Link® SMTP Interrater/Test Reliability System
Received: by cse.ucla.edu with SMTP;28 Apr 1998 11:32:18 -0700
Received: from IUSD.K12.CA.US (is_s1.iusd.k12.ca.us [172.16.111.252])
by cyrus.iusd.k12.ca.us (8.8.5/8.8.5) with SMTP id LAA11161
for <kim@cse.ucla.edu>; Tue, 28 Apr 1998 11:32:05 -0700 (PDT)
Received: from IUSDNET-Message_Server by IUSD.K12.CA.US
with Novell_GroupWise; Tue, 28 Apr 1998 11:31:48 -0700
Message-Id: <s545be24.0690@IUSD.K12.CA.US>
X-Mailer: Novell GroupWise 4.1
Date: Tue, 28 Apr 1998 11:31:14 -0700
From: Bev Huff <bhuff@iusd.k12.ca.us>
To: kim@cse.ucla.edu
Subject: Interrater/Test Reliability System
Mime-Version: 1.0
Content-Type: text/plain
Content-Disposition: inline

Kim, I recently purchased from you the ITRS. I received a disk and some documentation. I am having some problems importing data. Your documentation refers to another User Manual that might help me with the problem.

I have phoned both you and Jamal Abedi at the Advanced Data and Research Center. No one has returned my calls yet. I have an urgent need to get the problem solved soon, so what ever you can do to get me the User Manual would be appreciated.

Beverly Huff
Coordinator - Research, Evaluation and Assessment
Irvine Unified School District
bhuff@iusd.k12.ca.us

==========